CCN Assignment 2: A Simple Network Model of Perceptual Learning

Peggy Seriès, pseries@inf.ed.ac.uk

March 7, 2008

Practical

Deadline is March 24th, at noon (standard late policies apply). Hard-copies are preferred, but if you are out of town an email to me is okay (pdf or postscript format). Hand in to Pat Ferguson, Rm. D10, ANC, 5 Forrest Hill.

Report your findings, describe what you see and explain the results you get. Particularly well-researched answers can receive additional points. Ideally you substantiate your explanations for instance by additional simulations. Plots should include axes labels and units (either on the plot, or mentioned in the text), you can have a look at the webpage of Neural Computation (Mark Van Rossum's course) for more information about reports and graphics. You are marked on the quality of the report, not the code. Ideally, it should have the format and style of a scientific publication.

Copying results is not allowed. It's OK to ask for help from your friends. However, this help must not extend to copying code or written text that your friend has written, or that you and your friend have written together. I assess you on the basis of what you are able to do by yourself. It's OK to help a friend. However, this help must not extend to providing your friend with code or written text. If you are found to have done so, a penalty will be assessed against you as well.

Email me the Matlab script that you used, I will not assess the programming style, but I might check them if results are unexpected. I will also run plagiarism detectors on them. Email it to pseries@inf.ed.ac.uk and the subject should contain ccn2-2008 (all lowercase).

Aim of this assignment

The situation we are interested in is a two-interval discrimination task where on each trial 2 angles, $\theta_1 = \theta - \frac{\delta\theta}{2}$ and $\theta_2 = \theta + \frac{\delta\theta}{2}$ are shown in succession.

The task of the subject is to respond $+1$ if θ_1 preceded θ_2 and -1 otherwise. Experimental evidence shows that performance at such a discrimination task improves dramatically with practice. An interesting feature of this improvement is that it is stimulus selective. For example, it doesn't transfer well to directions other than the trained direction [2]. This limited transfer suggests that the learning is due to changes in early stages of the sensory pathway, where stimuli characterized by very different parameters are represented by different neurons. As the properties of the neurons in these early stages are relatively well known, especially in the visual cortex, we can attempt to use this information to study possible neural mechanisms of perceptual learning in these systems.

The aim of this assignment is to explore the properties of a very simple model of perceptual learning.

Note : questions 1.4, 1.5, and 1.7 are independent and can be treated in any order (1.6 is easier to address after 1.5).

1 Model

1.1 Model of the population of neurons (15 pts)

We consider a population of $N = 50$ neurons with tuning curves $f_i(\theta)$ describing the mean spike count of each neuron in 1 second as a function of the stimulus direction θ. The cells have preferred directions θ_i equally spaced between -180 deg and 180 deg. The tuning curves are circular normal distributions:

$$f(\theta) = G.\exp(\beta(\cos(\theta - \theta_i) - 1)) + K \tag{1}$$

where G is the maximal firing rate ($G = 50$ spk), $\beta = 3$ controls the width of the tuning curves and K denotes spontaneous activity ($K = 5$ spk). The variability of the spike count is Poisson. We denote by $\mathbf{r}(\theta) = \{r_1(\theta), .., r_N(\theta)\}$ the response of the population of neurons on a given trial of 1 sec.

1. On the same graph, plot an example of the population response $\mathbf{r}$ to stimulus $\theta_1 = -5^\circ$ and an example of the population response to stimulus $\theta_1 = +5^\circ$ for one trial.

1.2 Discrimination by a threshold linear network (25 pts)

We would like to understand what type of neural architecture is capable of performing the task. The simplest neural network that is capable of performing discrimination between two directions is a single-layer perceptron, which computes a linear sum of its inputs, followed by a thresholding. We consider the perceptron which takes as inputs the activities of the neurons modelled above, and has one output or 'decision' unit y whose activity on each trial t is described by [4]:

$$y^t = \text{sign}(R_1^t - R_2^t) \tag{2}$$

where:

$$R_1^t = \sum_{i=1}^{i=N} w_i r_i^t \tag{3}$$

$$R_2^t = \sum_{i=1}^{i=N} w_i {r'}_i^t \tag{4}$$

The output of the perceptron is thus +1 or -1. The w_i are the weights of the perceptron, and $\{r_i^t\}$ and $\{{r'}_i^t\}$ are the responses of the neurons to the first and second stimuli shown during the trial, respectively[1].

1. Initialize the weights $\{w_i\}$ to random numbers uniformely distributed between -10^{-6} and 10^{-6}. Plot those initial weights.

2. - We suppose that in the first half of the trials, θ_1 is presented in the first interval and θ_2 in the second. On the second half of the trials, θ_2 is presented first. In a matrix $X_{train,1}$ (N_txN), which will represent the responses of the neurons in the first interval, collect $\frac{N_t}{2}$ trials of population responses to θ_1 following by $\frac{N_t}{2}$ trials of population response to θ_2. Similarly, in a matrix $X_{train,2}$ representing the responses of the neurons in the second interval, collect $\frac{N_t}{2}$ trials of population responses to θ_2 followed by $\frac{N_t}{2}$ trials of population responses to θ_1. This is the *training* data. We use $N_t = 500$ trials. and $\theta_1 = -5$ deg and $\theta_2 = 5$ deg.
   - In a vector Y_{target}, store the corresponding desired output of the perceptron for all the training trials.
   - Similarly, in matrices $X_{test,1}$ (resp. $X_{test,2}$) collect $N_t/2$ trials of population responses to stimulus θ_1(resp. θ_2) followed by $N_t/2$ responses to θ_2 (resp. θ_1). This is the *testing* data.
   - To check that your data is correct, plot on the same graph the population response corresponding to the mean of $X_{train,1}$ averaged over the first $\frac{N_t}{2}$ trials, and the mean of $X_{train,1}$ averaged over the following $\frac{N_t}{2}$ trials.

3. Using the initial random weights, compute the output of the perceptron for all the training and testing trials (you can store the results in Y_{train} and Y_{test}), and report the percentage of correct discrimination.

[1] This model presupposes the existence of a short term memory that stores the summed response R_1 of the first stimulus, before it is compared to R_2 the summed response to the second stimulus. On some trials, θ_1 is presented first; on other trials, θ_2.

1.3 Supervised Learning of the Weights (30 pts)

Perceptual learning is modelled as an optimization of the weights of the perceptron for the discrimination task. Learning of the weights is based on gradient descent, using the 'delta rule' [1]: after each trial, the output y_{train} (± 1) is compared to the correct answer y_{target} (± 1); If the difference is zero, no learning takes place; otherwise, the weights are adjusted to reduce this difference. The change in weight from each neuron of the network to the decision unit is given by:

$$w_i = w_i + \eta(r_i^t - {r'}_i^t)(y_{target}^t - y_{train}^t) \tag{5}$$

where η is the learning rate, r_i^t is the response of neuron i for the first stimulus, and $r_i'^t$ is the response of neuron i for the second stimulus, y_{target}^t is the desired output and y_{train}^t is the actual output of the perceptron[2][3].

1. Define a stopping criterion for the learning: possible (non-exclusive) choices are i) stop learning when the error $E = \sum_t (y_{target}^t - y_{train}^t)^2 / N_t$ is lower than a given threshold; ii) stop learning when error on the testing set increases (to prevent overfitting); iii) stop learning when the number of iterations exceeds some maximum number (e.g. 1000 presentations of all the data).
2. Set $\eta = 10^{-8}$. Implement the learning rule and check that the error E decreases during learning.
3. Train the perceptron a few times (i.e. with different initial weight configurations). On average, how many iterations are necessary for the network to reach the stopping criterion?
4. Change the value of the learning rate η and comment.
5. Plot the resulting optimal weights for a representative 'best' run, and report the classification correct after learning. Can you explain the shape of the optimal weights?
6. Learn again for $\theta_{1,2} = \pm 1$ deg (fine discrimination), ± 45 deg, ± 90 deg (broad discrimination). How do the weights change?

[2] We use upper case letters to denote matrices and vectors and lower case to denote scalars. y_{target}^t is the t^{th}element of Y_{target} above.

[3] This description of learning corresponds to *online* learning (the update of the weights is made after presentation of each stimulus pairs). You can also use *batch* learning where the whole batch of examples is used before a single update is made. In that case, the learning rule is:

$$w_i = w_i + \eta \sum_t (r_i^t - {r'}_i^t)(y_{target}^t - y_{train}^t) \tag{6}$$

1.4 Network Performance after Learning (15 pts)

1. What is approximately the discrimination threshold, or the smallest angle difference that can be reliably discriminated on 74% of the trials ? (note: decreasing the learning rate can help resolving smaller angles).
2. What would be the corresponding performance of an optimal read-out ? (hint: write Fisher Information and use its relation with the discrimination threshold).
3. What other aspects of the model could then be changed during perceptual learning which would further improve discriminability?

1.5 Transfer of learning (15 pts)

1. Learn the weights to discriminate $\theta_1 = -2$ deg and $\theta_2 = 2$ deg (fine discrimination). Describe the performance of the network for discriminating other angles separated by the same amount, e.g. $\{0, 4\}, \{4, 8\}$, $\{8, 12\}$, etc..
2. Describe the performance of the network for discriminating broader angles, e.g. ± 3 deg, ± 4 deg, ± 5 deg etc...
3. Based on these results, what can you say about how learning transfers to untrained directions with this model?

1.6 Bonus points: Deterioration (10 pts)

In psychophysics, some studies have found that learning of a certain task (A) can be disrupted or interfered with by subsequent practice of a second task (B) (see e.g.[3]). For example, if on each day of the practice sessions, subjects are asked to train on both tasks A and B, for some configurations of the tasks, they end up not learning A nor B at all. On the contrary, if they are trained uniquely on A or B, they learn smoothly.

1. By alternating learning between different pairs of angles, could we get a better transfer with this model ? or are performances going to deteriorate for all angles?

1.7 Bonus Points: Application to a Detection Task (10 pts)

1. The same framework can be used to model a detection task. Explain how.

2. What would be the shape of the weights after learning to detect a direction of θ_0 (e.g. 0 deg)? (hint: decrease the gain of the stimulus evoked G to about 3 spk to make it a realistic detection task).

3. Task transfer: If the model is trained to detect a direction of θ_0, does it become better at discriminating around θ_0 ?

2 Bonus Point: Conclusion (5 pts)

To what extent do you think this is an interesting model of perceptual learning? How could we improve it?

References

[1] Peter Dayan and Larry F Abbott. *Theoretical Neuroscience*. MIT Press, 2001.

[2] Manfred Fahle. Perceptual learning: specificity versus generalization. *Curr Opin Neurobiol*, 15(2):154–160, Apr 2005.

[3] Aaron R Seitz, Noriko Yamagishi, Birgit Werner, Naokazu Goda, Mitsuo Kawato, and Takeo Watanabe. Task-specific disruption of perceptual learning. *Proc Natl Acad Sci U S A*, 102(41):14895–14900, Oct 2005.

[4] H.S. Seung and Sompolinsky H. A simple model for reading neuronal population codes. 90:10749–10753, 1993.